



PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of

Hans-Jurgen MATT, et al.

Attorney Docket Q61703

Appln. No.: 09/716,272

Group Art Unit: 2681

Technology Center 2600

Confirmation No.: not yet assigned

Examiner: Not yet assigned

Filed: November 21, 2000

For: EXPONENTIAL ECHO AND NOISE REDUCTION IN SILENCE INTERVALS


SUBMISSION OF PRIORITY DOCUMENT

Commissioner for Patents  
Washington, D.C. 20231

Sir:

Submitted herewith is a certified copy of the priority document on which a claim to priority was made under 35 U.S.C. § 119. The Examiner is respectfully requested to acknowledge receipt of said priority document.

Respectfully submitted,

  
David J. Cushing  
Registration No. 28,703

SUGHRUE, MION, ZINN,  
MACPEAK & SEAS, PLLC  
2100 Pennsylvania Avenue, N.W.  
Washington, D.C. 20037-3213  
Telephone: (202) 293-7060  
Facsimile: (202) 293-7860

Enclosures: Certified Copy of German Patent Application No. 19957221.6

Date: April 19, 2001

Best Available Copy

#4  
4/27/01  
MB  
RECEIVED

APR 20 2001

# BUNDESREPUBLIK DEUTSCHLAND



09/716,272  
Q61703  
181

## Prioritätsbescheinigung über die Einreichung einer Patentanmeldung

**Aktenzeichen:** 199 57 221.6

**Anmeldetag:** 27. November 1999

**Anmelder/Inhaber:** Alcatel, Paris/FR

**Bezeichnung:** Exponentielle Echo- und Geräuschabsenkung in Sprachpausen

**IPC:** G 10 L, H 04 M

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ursprünglichen Unterlagen dieser Patentanmeldung.

München, den 21. November 2000  
Deutsches Patent- und Markenamt  
Der Präsident  
Im Auftrag

*Sammer*

**CERTIFIED COPY OF  
PRIORITY DOCUMENT**

## Patentansprüche

1. Verfahren zur Reduktion von Echo- und/oder Geräuschsignalen bei Telekommunikations(=TK)-Systemen für die Übertragung von akustischen Nutzsignalen, insbesondere menschlicher Sprache, bei dem mittels Sprach-Pausen-Detektion ermittelt wird, wann in der zu übertragenden Mischung aus Nutzsignalen und Störsignalen ein Sprachsignal enthalten ist oder wann eine Sprachpause vorliegt, wobei mittels eines Multiplizierers mit zwei Eingängen die in der Regel durch Echo- und/oder Geräuschsignale gestörten Nutzsignale in ihrer Amplitude durch ein zeitabhängiges Steuersignal  $a_0(t)$  bzw. durch ein im Rhythmus einer Abtastrate  $f_T=1/T$  getaktetes Steuersignal  $a_0(k)$  verändert werden, wobei  $k \in \mathbb{N}$  die Abtastwerte durchzählt und  $T$  die Periodendauer von einem Abtastwert zum nächsten bedeutet,

dadurch gekennzeichnet,

dass das Steuersignal  $a_0(t)$  bzw.  $a_0(k)$  so variiert wird, dass während des Vorliegens von Sprachsignalen im Nutzsignal die Amplitude des Steuersignals  $a_0(t)$  bzw.  $a_0(k)$  auf einen vorgegebenen konstanten Wert  $c_0$  gesetzt wird und mit Beginn einer Sprachpause im Nutzsignal die Amplitude des Steuersignals  $a_0(t)$  bzw.  $a_0(k)$  von einem Abtastwert zum nächsten gemäß der Rekursionsformel

$$a_0(k+1) = a_0(k) \cdot \beta \quad \text{mit } \beta < 1$$

stetig abgesenkt wird,

und dass nach dem Ende einer Sprachpause  $a_0(k) = c_0$  gesetzt wird.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass der Faktor  $\beta$  aus der Abtastrate  $f_T$ , aus einer Zeitkonstanten  $\tau_1$  und aus einem vorgegebenen konstanten Vorfaktor  $c_1$  bestimmt wird gemäß der Beziehung
 
$$\beta = c_1 \cdot \exp(-1/\tau_1 f_T) .$$
3. Verfahren nach Anspruch 2, dadurch gekennzeichnet, dass die Zeitkonstante  $\tau_1$  zwischen 50 ms und 150 ms gewählt wird, vorzugsweise  $\tau_1 \approx 65$  ms.
4. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass der konstante Wert  $c_0 = 1$  gewählt wird.
5. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass  $a_0(k+1)$  während einer Sprachpause und/oder des Vorliegens eines Echsignals einen vorgegebenen konstanten Wert  $c_2$  annimmt, falls der Vorgängerwert  $a_0(k) \leq c_2$  geworden ist.
6. Verfahren nach einem der Ansprüche 1 bis 4, dadurch gekennzeichnet, dass während einer Sprachpause und/oder des Vorliegens eines Echsignals und für  $a_0(k) \leq c_2$ , wobei  $c_2$  eine vorgegebene Konstante ist, der Leistungswert des Geräuschpegels  $N$  im aktuell benutzten TK-Kanal fortlaufend gemessen und/oder geschätzt wird, und dass in Abhängigkeit vom aktuellen Geräuschpegel  $N$  laufend das Steuersignal  $a_0(k+1)$  eingestellt wird gemäß  $a_0(k+1) = f(N)$ , wobei  $f(N)$  eine vorgegebene Funktion von  $N$  ist.
7. Verfahren nach Anspruch 6, dadurch gekennzeichnet, dass die vorgegebene Funktion  $f(N)$  eine Funktion  $g(S/N)$  ist, die vom Quotienten  $S/N$  aus dem Leistungswert des Signalpegels  $S$  der zu übertragenden Nutzsignale und dem Leistungswert des

Geräuschpegels  $N$  abhängt, oder dass die vorgegebene Funktion  $f(N)$  eine Funktion  $g'(N/S)$  ist, die vom Kehrwert  $N/S$  dieses Quotienten abhängt.

8. Verfahren nach Anspruch 7, dadurch gekennzeichnet, dass die Funktion  $f(N)$  bzw.  $g(S/N)$  bei  $1/N \ll 1$  bzw.  $S/N = 0$  dB mit einem konstanten Wert  $f_0 > 0$  bzw.  $g_0 > 0$  beginnt, im Bereich zwischen  $N$  bzw.  $S/N = 10$  dB bis 15 dB, vorzugsweise bei  $N$  bzw.  $S/N \approx 12$  dB, auf ein Maximum  $f_{\max}$  bzw.  $g_{\max}$  ansteigt und anschließend auf einen Minimalwert  $f_{\min}$  bzw.  $g_{\min}$ , vorzugsweise auf 0 dB abfällt, wobei  $5 \text{ dB} \leq f_0, g_0 \leq 10 \text{ dB}$ , vorzugsweise  $6 \text{ dB} \leq f_0, g_0 \leq 8 \text{ dB}$ , und wobei  $20 \text{ dB} \leq f_{\max}, g_{\max} \leq 30 \text{ dB}$ , vorzugsweise  $f_{\max}, g_{\max} \approx 25 \text{ dB}$ .
9. Verfahren nach einem der Ansprüche 6 bis 8, dadurch gekennzeichnet, dass die Funktion  $f(N)$  bzw.  $g(S/N)$  zumindest stückweise, vorzugsweise in allen Teilabschnitten linear mit  $N$  bzw.  $S/N$  verläuft.
10. Verfahren nach einem der Ansprüche 6 bis 8, dadurch gekennzeichnet, dass die Funktion  $f(N)$  bzw.  $g(S/N)$  aus Polynomen aufgebaut ist und als unsymmetrische Glockenkurve über  $N$  bzw.  $S/N$  verläuft.
11. Verfahren nach einem der Ansprüche 6 bis 10, dadurch gekennzeichnet, dass die Funktionen  $f(N)$  bzw.  $g(S/N)$  oder  $g'(N/S)$  so gewählt werden, dass die Reduktion des Geräuschpegels  $N$  gehör richtig gemäß den psychoakustischen Mittelwerten des menschlichen Gehörspektrums erfolgt.

12. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass zusätzlich zur Erkennung und Reduktion von Geräuschsignalen das Vorliegen von Echosignalen detektiert und/oder vorhergesagt wird und die Echosignale unterdrückt oder reduziert werden.
13. Verfahren nach Anspruch 12 und einem der Ansprüche 6 bis 11, dadurch gekennzeichnet, dass das Steuersignal  $a_0(k+1)$  laufend eingestellt wird gemäß  $a_0(k+1) = h(N, S, ES, \tau_E, ERL)$ , wobei  $h(N, S, ES, \tau_E, ERL)$  eine vorgegebene Funktion von  $N$ ,  $S$ , dem Nutzsignal  $ES$  in Gegenrichtung eines sprechenden TK-Partners,  $\tau_E$  eine konstante Verzögerungszeit des Echosignals und  $ERL$  eine Dämpfungskonstante der Amplitude des Echosignals ist.
14. Verfahren nach Anspruch 12, dadurch gekennzeichnet, dass die Steuerung der Reduktion von Geräuschsignalen und der Reduktion von Echosignalen getrennt erfolgt.
15. Verfahren nach einem der Ansprüche 12 bis 14, dadurch gekennzeichnet, dass während der Zeitdauer einer Echo-Reduktion zum Nutzsignal zusätzlich ein künstliches Geräuschsignal addiert wird.
16. Verfahren nach Anspruch 15, dadurch gekennzeichnet, dass das künstliche Geräuschsignal eine psychoakustisch als angenehm empfundene akustische Signalsequenz (=comfort noise) umfasst.
17. Verfahren nach Anspruch 15, dadurch gekennzeichnet, dass das künstliche Geräuschsignal ein zuvor während der aktuellen TK-Verbindung aufgezeichnetes Geräuschsignal umfasst.

18. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet,  
dass in einem Sprach-Pausen-Detektor (SPD) aus dem Eingangssignal  $x$  mittels eines Kurzzeit-Pegelschätzers ein Kurzzeit-Ausgangssignal  $sam(x)$ , mittels eines Mittelzeit-Pegelschätzers ein Mittelzeit-Ausgangssignal  $mam(x)$  und mittels eines Langzeit-Pegelschätzers ein Langzeit-Ausgangssignal  $lam(x)$  gebildet wird,  
dass die drei Ausgangssignale  $sam(x)$ ,  $mam(x)$  und  $lam(x)$  über geeignete Verstärkungskoeffizienten so eingestellt werden, dass sie etwa gleich groß, wenn das Eingangssignal  $x$  ein reines Rauschsignal ist, wobei  $sam(x) < mam(x) < lam(x)$ ,  
dass die drei Ausgangssignale  $sam(x)$ ,  $mam(x)$  und  $lam(x)$  von Komparatoren überwacht werden,  
und dass das Vorliegen eines Sprachsignals als Eingangssignal  $x$  angenommen wird, wenn  $sam(x)$  und  $mam(x)$  zunächst jeweils größer werden als  $lam(x)$ , und das Vorliegen einer Sprachpause, wenn danach  $sam(x)$  und/oder  $mam(x)$  wieder kleiner wird als  $lam(x)$ .
19. Verfahren nach Anspruch 18, dadurch gekennzeichnet, dass die drei Ausgangssignale  $sam(x)$ ,  $mam(x)$  und  $lam(x)$  zur Sprach-Pausen-Abschätzung einem neuronalen Netz zugeführt werden, das mit einer Vielzahl von Szenarien mit unterschiedlichen Eingangssignalen  $x$  trainiert wurde.
20. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass das zu übertragende Nutzsignal einer spektralen Subtraktion unterzogen wird.
21. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass das zu übertragende Nutzsignal einer dem

menschlichen Gehör angepassten spektralen Filterung unterzogen wird.

22. Servereinheit zur Unterstützung des Verfahrens nach einem der Ansprüche 1 bis 21.

23. Computerprogramm zur Durchführung des Verfahrens nach einem der Ansprüche 1 bis 21.



Anmelder:

Alcatel  
54, Rue La Boétie  
75088 Paris  
Frankreich

Stuttgart, den 26.11.1999

V1267-140B B/ho

~~111466B~~

111 705

Vertreter:

Kohler Schmid + Partner  
Patentanwälte GbR  
Ruppmannstraße 27  
70565 Stuttgart

Exponentielle Echo- und Geräuschabsenkung in Sprachpausen

Die Erfindung betrifft ein Verfahren zur Reduktion von Echo- und/oder Geräuschsignalen bei Telekommunikations(=TK)-Systemen für die Übertragung von akustischen Nutzsignalen, insbesondere menschlicher Sprache, bei dem mittels Sprach-Pausen-Detektion ermittelt wird, wann in der zu übertragenden Mischung aus Nutzsignalen und Störsignalen ein Sprachsignal enthalten ist oder wann eine Sprachpause vorliegt, wobei mittels eines Multiplizierers mit zwei Eingängen die in der Regel durch Echo- und/oder Geräuschsignale gestörten Nutzsignale in ihrer Amplitude durch ein zeitabhängiges Steuersignal  $a_0(t)$  bzw. durch ein im Rhythmus einer Abtastrate  $f_T = 1/T$  getaktetes Steuersignal  $a_0(k)$  verändert werden, wobei  $k \in \mathbb{N}$  die Abtastwerte durchzählt und  $T$  die Periodendauer von einem Abtastwert zum nächsten bedeutet.

Ein solches Verfahren ist beispielsweise bekannt aus der DE 42 29 912 A1.

Während einer natürlichen Kommunikation zwischen Menschen passt man in der Regel die Amplitude der gesprochenen Sprache automatisch an die akustische Umgebung an. Bei einer Sprachkommunikation zwischen entfernten Orten jedoch befinden sich die Gesprächspartner nicht im selben akustischen Umfeld und sind sich daher jeweils nicht der akustischen Situation am Ort des anderen Gesprächspartners bewusst. Besonders verschärft tritt daher ein Problem auf, wenn einer der Partner aufgrund seiner akustischen Umgebung gezwungen ist, sehr laut zu sprechen, während der andere Partner in einer leisen akustischen Umgebung Sprachsignale mit geringer Amplitude erzeugt.

Hinzu kommt das Problem, dass auf einem TK-Kanal auch ein "elektronisch erzeugtes" Geräusch entsteht und als Hintergrund zum Nutzsignal mitübertragen wird. Des weiteren ist es auch vorteilhaft, Störsignale wie unerwünschten Hintergrundlärm (Straßenlärm, Fabriklärm, Bürolärm, Kantinenlärm, Fluglärm etc.) zu reduzieren oder ganz zu unterdrücken. Um den Komfort beim Telefonieren zu erhöhen, ist man generell bestrebt, jede Art von Geräusch möglichst gering zu halten.

Schließlich entstehen bei TK-Verbindungen auch noch sogenannte Echos, die in Zweidraht-TK-Netzen als Leitungsechos und beispielsweise in einfachen und unkomfortableren TK-Endgeräten in Form von akustischen Echos auftreten.

Allgemein ist es deshalb bei der Übertragung von einem Gemisch aus Sprachsignal und Störsignalen über TK-Netze wichtig, die Störsignale wie Geräusch und Echo soweit wie möglich in ihrer Amplitude abzusenken.

Ein bekanntes Verfahren zur Geräuschreduktion ist die sogenannte "spektrale Subtraktion", die beispielsweise in der Veröffentlichung "A new approach to noise reduction based on auditory masking effects" von S. Gustafsson und P.

Jax, ITG-Fachtagung, Dresden, 1998, beschrieben ist. Dabei handelt es sich um ein spektrales Geräuschabsenkungsverfahren, bei dem eine akustische Verdeckungsschwelle (beispielsweise nach dem MPEG-Standard) berücksichtigt wird. Nachteilig bei derartigen Verfahren ist die aufwendige Bestimmung dieser akustischen Verdeckungsschwelle und die Ausführung aller mit diesem Verfahren verbundenen Rechenoperationen.

Bei einer spektralen Subtraktion wird zunächst das Geräusch in den Sprachpausen gemessen und in Form eines Leistungsdichtespektrums fortlaufend in einem Speicher abgelegt. Das Leistungsdichtespektrum wird über eine Fourier Transformation gewonnen. Beim Auftreten von Sprache wird sodann das gespeicherte Geräuschspektrum "als bester aktueller Schätzwert" vom aktuellen gestörten Sprachspektrum subtrahiert, sodann in den Zeitbereich zurücktransformiert, um auf diese Weise eine Geräuschreduktion für das gestörte Signal zu erhalten.

Ein weiterer Nachteil der spektralen Subtraktion besteht darin, dass durch den Vorgang einer prinzipiell nicht genauen spektralen Geräuschschätzung und nachfolgender Subtraktion auch Fehler im Ausgangssignal auftreten, die sich als "musical tones" bemerkbar machen. Außerdem ist dieses bekannte Verfahren kaum zur Unterdrückung von Echosignalen bei TK-Verbindungen geeignet.

Bei der erweiterten spektralen Signalbearbeitung, die ebenfalls in dem genannten Zitat beschrieben ist, werden zunächst mit Hilfe einer spektralen Subtraktion die Leistungsdichtespektren für das Geräusch und für die Sprache selbst geschätzt. Aus der Kenntnis dieser Teilspektren wird sodann mit Hilfe z.B. der Regeln aus dem MPEG Standard eine spektrale akustische Verdeckungsschwelle  $R_T(f)$  für das menschliche Ohr berechnet. Mit Hilfe dieser Verdeckungsschwelle und den geschätzten Spektren für Geräusch und Sprache wird sodann nach einer einfachen Regel eine Filterdurchlasskurve  $H(f)$  berechnet, die so gestaltet ist, dass wesentliche spektrale Teile der Sprache

möglichst unverändert durchgelassen und spektrale Teile des Geräusches möglichst abgesenkt werden.

Sodann wird das originale gestörte Sprachsignal nur durch dieses Filter gegeben, um auf diese Weise eine Geräuschreduktion für das gestörte Signal zu erhalten. Der Vorteil dieses Verfahrens besteht nun darin, dass vom gestörten Signal "Nichts addiert oder subtrahiert" wird und daher Fehler in den Schätzungen weniger bis kaum wahrnehmbar sind. Nachteilig ist wieder der erhebliche Rechenaufwand für die spektrale Geräuschunterdrückung sowie das für eine Echounterdrückung vorzuschaltende adaptive Filter.

Bei dem bekannten Kompander-Verfahren, wie es beispielsweise in der eingangs zitierten DE 42 29 912 A1 beschrieben ist, wird der Grad der Geräusch- und Echoabsenkung gemäß einer fest vorgegebenen Transferfunktion festgelegt, die unter anderem eine Pegelabsenkung auch bei sehr kleinen Eingangssignalen vornimmt.

Der Kompander hat zunächst die Eigenschaft, Sprachsignale mit einem bestimmten (vorab eingestellten) "normalen Sprachsignalpegel" (ggf. normale Lautstärke genannt) praktisch unverändert von seinem Eingang zum Ausgang zu übertragen.

Wird nun aber das Eingangssignal einmal zu laut, z.B. weil ein Sprecher zu dicht an sein Mikrofon kommt, so begrenzt ein Dynamik-Kompressor den Ausgangspegel auf nahezu den gleichen Wert wie im Normalfall, indem die aktuelle Verstärkung im Kompander mit zunehmender Eingangslautstärke linear abgesenkt wird. Durch diese Eigenschaft bleibt die Sprache am Ausgang des Kompandersystems etwa gleich laut – unabhängig davon wie stark die Eingangslautstärke schwankt.

Wird nun andererseits ein Signal mit einem Pegel, der kleiner als der Normalpegel ist, auf den Eingang des Kompanders gegeben, so wird das

Signal zusätzlich gedämpft, indem die Verstärkung zurückgeregelt wird, um Hintergrundgeräusche möglichst nur abgeschwächt zu übertragen.

Der Kompander besteht somit aus einem Kompressor für Sprachsignalpegel, die größer oder gleich einem Normalpegel sind und einem Expander für Signalpegel, die kleiner als der Normalpegel sind. Die Verstärkungsabsenkung im Expander wird dabei mit zunehmend kleineren Eingangspegeln stärker.

Nachteilig bei der Kompander-Lösung ist der erhebliche Rechenaufwand, der zur Durchführung des bekannten Verfahrens erforderlich ist. Durch die Kompression des Sprachsignalpegels einerseits und durch die Expansion andererseits wird außerdem eine Modulation in der Sprachlautstärke hervorgerufen, die das Sprachsignal in einer Weise verändert, dass das Ergebnis subjektiv oft als unbefriedigend empfunden wird, d.h. einen unbefriedigenden Höreindruck hinterlässt.

Aufgabe der vorliegenden Erfindung ist es demgegenüber, ein Verfahren mit den eingangs beschriebenen Merkmalen vorzustellen, bei dem in möglichst unaufwendiger und kostengünstiger Art und Weise ohne großen Rechenaufwand und mit geringem Bedarf an Rechenspeicher und Datenspeicherplatz eine Echo- und Geräuschabsenkung bewirkt wird, die mit einfachen Mitteln einen für das menschliche Ohr möglichst angenehmen akustischen Gesamteindruck erzeugt, der je nach Geschmack zusätzlich an individuelle Bedürfnisse angepasst werden kann.

Erfindungsgemäß wird diese Aufgabe auf ebenso einfache wie wirkungsvolle Art und Weise dadurch gelöst, dass das Steuersignal  $a_0(t)$  bzw.  $a_0(k)$  so variiert wird, dass während des Vorliegens von Sprachsignalen im Nutzsignal die Amplitude des Steuersignals  $a_0(t)$  bzw.  $a_0(k)$  auf einen vorgegebenen konstanten Verstärkungswert  $c_0$  gesetzt wird und mit Beginn einer Sprachpause im Nutzsignal die Amplitude des Steuersignals  $a_0(t)$  bzw.  $a_0(k)$  von einem Abtastwert zum nächsten gemäß der Rekursionsformel

$$a_0(k+1) = a_0(k) \cdot \beta \quad \text{mit } \beta < 1$$

stetig abgesenkt wird,

und dass nach dem Ende einer Sprachpause wieder  $a_0(k) = c_0$  gesetzt wird.

Damit steht eine sehr einfache und sehr kostengünstige Methode bereit, die auch eine überraschend gute Qualität bezüglich einer Störreduktion ergibt, indem sie vorzugsweise in Sprachpausen die störenden Echo- und Geräuschsignale reduziert. Während der Sprachphasen selbst werden die Störgeräusche zumindest teilweise maskiert und daher vom menschlichen Ohr weit weniger deutlich wahrgenommen. Durch das Weglassen der Kompression nach dem bekannten Kompander-Verfahren wird das originale Sprachsignal erheblich weniger verändert, so dass im Ergebnis ein in der Regel besser klingendes Sprachsignal am anderen Ende der Leitung ankommt. Außerdem erfordert das erfindungsgemäße Verfahren eine geringere Rechenleistung als das Kompander-Verfahren, da ja zumindest die Kompression unterbleibt. Entsprechend sind geringere Kapazitäten an Datenspeichern und Rechenspeicher erforderlich, was das erfindungsgemäße Verfahren im Gegensatz zu den bekannten Verfahren einfacher und kostengünstiger gestaltet.

Um eine effektive Geräuschreduktion zu erzielen, wird das zu übertragende Signal während der Sprachpausen in seinem Leistungswert gemäß einer zeitlichen Exponentialfunktion, im Gegensatz zu einer vom Eingangspegel abhängigen Absenkung, wie bei dem Kompander-Verfahren, abgesenkt. Damit wird bereits eine wesentliche Geräuschminderung erzielt. Hinzu kommt, dass eine Absenkung der Geräusche während einer Sprachpause das Gehör deutlich weniger belastet, indem es den Taubheitseffekt nach lauter Schalleinwirkung wesentlich mindert. Das Ohr kann bei Wiedereinsetzen der Sprache empfindlicher reagieren und genauer hinhören.

Beispielsweise wird man Geräusche vorzugsweise in Abhängigkeit vom momentanen Geräuschpegel  $N$  bzw. in Abhängigkeit von einer Funktion  $g(S/N)$  des Signal-zu-Geräuschabstandes  $S/N$  absenken, aber kurzzeitig auftretende Echos stärker absenken und nach dem Ende der Echos die Absenkung auf den geringeren Wert der Geräuschabsenkung zurückführen.

Besonders bevorzugt ist daher eine Verfahrensvariante, die sich dadurch auszeichnet, dass bei Vorliegen eines Geräuschsignals und/oder Echosignals und für  $a_0(k) \leq c_2$ , wobei  $c_2$  eine vorgegebene Konstante ist, der Leistungswert des Geräuschpegels  $N$  im aktuell benutzten TK-Kanal fortlaufend gemessen und/oder geschätzt wird, und dass in Abhängigkeit vom aktuellen Geräuschpegel  $N$  laufend das Steuersignal  $a_0(k+1)$  eingestellt wird gemäß  $a_0(k+1) = f(N)$ , wobei  $f(N)$  eine vorgegebene Funktion von  $N$  ist.

Damit wird der Grad einer Geräuschabsenkung vom aktuell auftretenden Leistungswert  $N$  des Geräusches automatisch mitgesteuert und dem aktuellen Geräuschwert im Telefonkanal angepasst und in vorbestimmter definierter Weise nachgeführt. Über die Wahl der Funktion  $f(N)$  kann auch der subjektive Eindruck des erzeugten Gesamtsignals angepasst werden. Ein weiterer Vorteil dieser Verfahrensvariante besteht darin, dass bei einem Bündel von Telefonkanälen, beispielsweise zwischen internationalen Vermittlungsstellen, die Geräuschsituation in jedem einzelnen Kanal, die ja von Kanal zu Kanal sehr verschieden sein kann, automatisch eingestellt und individuell optimiert werden kann.

Besonders bevorzugt ist eine Variante des erfindungsgemäßen Verfahrens, die sich dadurch auszeichnet, dass die vorgegebene Funktion  $f(N)$  eine Funktion  $g(S/N)$  ist, die vom Quotienten  $S/N$  aus dem Leistungswert des Signalpegels  $S$  der zu übertragenden Nutzsignale und dem Leistungswert des Geräuschpegels  $N$  abhängt, oder dass die vorgegebene Funktion  $f(N)$  eine Funktion  $g'(N/S)$  ist, die vom Kehrwert  $N/S$  dieses Quotienten abhängt. Aus Gründen einer

einfacheren praktischen Realisierung kann man auch eine Funktion von  $(S+N)/N$  oder von  $(S+N)/S$  verwenden.

Der Vorteil der obigen Verfahrensvariante besteht darin, dass bei stark variierendem Nutzsignalpegel  $S$  in den Telefonkanälen eines Bündels immer die richtige Einstellung für die Geräuschabsenkung gefunden wird. Bei einer Steuerung der Geräuschreduktion proportional zum Kehrwert  $N/S$  lässt sich die Funktion  $g'(N/S)$  leicht auf einem digitalen Signalprozessor (=DSP) mit festen Computerwortlängen von beispielsweise 16 bit unter Verwendung von besonders einfacher Software implementieren, da für  $N/S$  vorzugsweise ein Zahlenbereich  $0 < N/S < 1$  zur Steuerung der Geräuschreduktion relevant bzw. interessant ist.

Akustische Gehörtests haben ergeben, dass bei  $S/N = 0$  dB die Sprache bereits so stark gestört ist, dass man das Geräusch nur bedingt um einen Wert  $f_0$  bzw.  $g_0$  zwischen 5 und 10 dB, vorzugsweise zwischen 6 und 8 dB reduzieren darf, um den akustischen Gesamteindruck im Hinblick auf eine Natürlichkeit der Sprache nicht zu verschlechtern. Bei noch ungünstigeren Werten des Signal-zu-Rausch-Verhältnisses  $S/N < 0$  dB kann dann der Wert  $f_0$  bzw.  $g_0$  beibehalten werden, da jede weitere Geräuschabsenkung den Gesamteindruck nur verschlechtert.

Bei mittleren  $S/N$  kann gemäß diesen Untersuchungen eine stärkere Geräuschabsenkung vorgenommen werden. Ein Maximum ergibt sich dabei im Bereich 10 bis 15 dB. Der Wert der Geräuschabsenkung  $f_{\max}$  bzw.  $g_{\max}$  sollte im Maximum zwischen 20 und 30, vorzugsweise etwa 25 dB betragen.

Bei sehr guten Rauschwerten  $S/N > 40$  dB sollte nur noch eine minimale Absenkung zwischen 0 und 3 dB eingestellt werden, um die Natürlichkeit der übertragenen Sprache so gut wie möglich zu erhalten.



Der Klang der Sprache und die Verständlichkeit sind besonders gut, wenn die Funktion  $f(N)$  bzw.  $g(S/N)$  über die drei oben diskutierten Bereiche hinweg in stetiger Weise miteinander zusammenhängt, wobei schnelle Änderungen in  $N$  oder in  $S(N)$  vorteilhaft durch Filterungen geglättet werden können.

Eine relativ einfache Realisierung in Hardware und/oder Software ergibt sich, indem man die Funktionen  $f(N)$  bzw.  $g(S/N)$  bzw.  $g'(N/S)$  durch gerade Kennlinienstücke zwischen den drei oben beschriebenen Betriebspunkten approximiert (abschnittsweise lineare Näherung).

Bei einer etwas aufwendigeren Variante des erfindungsgemäßen Verfahrens, die aber im Ergebnis zu einem besseren Klangbild führt, wird eine Polynomfunktion zur Implementierung der stetigen Funktionen  $f(N)$  bzw.  $g(S/N)$  bzw.  $g'(N/S)$  in den drei diskutierten Bereichen herangezogen, was im Ergebnis zu einer Art unsymmetrischer Glockenfunktion führt.

Besonders bevorzugt ist eine Variante des erfindungsgemäßen Verfahrens, bei der die Funktionen  $f(N)$  bzw.  $g(S/N)$  oder  $g'(N/S)$  so gewählt werden, dass die Reduktion des Geräuschpegels  $N$  gehörrichtig gemäß den psychoakustischen Mittelwerten des menschlichen Gehörspektrums erfolgt. Dabei wird der Wert für  $S$  und/oder  $N$  nicht nur aus dem momentanen Leistungswert alleine, sondern auch aus einem gewichteten spektralen Verlauf von  $S$  bzw.  $N$  bestimmt und insgesamt über die so gewonnene Funktion eine gehörrichtige, d.h. eine psychoakustisch angenehm klingende Geräuschreduktion erzielt. Da es kein einfach darstellbares Maß für eine akustisch angenehm klingende Geräuschreduktion gibt, sind alle Qualitätsbeurteilungen auf umfangreiche Gehörtests angewiesen, die anschließend mittels dafür optimierter statistischer Methoden ausgewertet werden, um einen Bewertungsmaßstab, (ähnlich wie bei Sprachcodecs) zu erhalten.

Eine gute Geräuschpegel-Schätzung erfordert einen guten Sprach-Pausen-Detektor, da man nur dann sicher sein kann, dass in den Sprachpausen-

Abschnitten lediglich störendes Geräusch und nicht irgendeine Mischung zwischen Geräusch und Sprachfetzen vorliegt, wie es in der Praxis häufig vorkommt.

Besonders bevorzugt ist daher eine Verfahrensvariante, die sich dadurch auszeichnet, dass im Sprach-Pausen-Detektor aus dem Eingangssignal  $x$  mittels eines Kurzzeit-Pegelschätzers ein Kurzzeit-Ausgangssignal  $\text{sam}(x)$ , mittels eines Mittelzeit-Pegelschätzers ein Mittelzeit-Ausgangssignal  $\text{mam}(x)$  und mittels eines Langzeit-Pegelschätzers ein Langzeit-Ausgangssignal  $\text{lam}(x)$  gebildet wird, dass die drei Ausgangssignale  $\text{sam}(x)$ ,  $\text{mam}(x)$  und  $\text{lam}(x)$  über geeignete Verstärkungskoeffizienten so eingestellt werden, dass sie etwa gleich groß, wenn das Eingangssignal  $x$  ein reines Rauschsignal ist, wobei  $\text{sam}(x) < \text{mam}(x) < \text{lam}(x)$ , dass die drei Ausgangssignale  $\text{sam}(x)$ ,  $\text{mam}(x)$  und  $\text{lam}(x)$  von Komparatoren überwacht werden, und dass das Vorliegen eines Sprachsignals als Eingangssignal  $x$  angenommen wird, wenn  $\text{sam}(x)$  und  $\text{mam}(x)$  zunächst jeweils größer werden als  $\text{lam}(x)$ , und das Vorliegen einer Sprachpause, wenn danach  $\text{sam}(x)$  und/oder  $\text{mam}(x)$  wieder kleiner wird als  $\text{lam}(x)$ .

Mit Hilfe dieser relativ einfachen Arten der Bildung von verschiedenen Mittelwerten des Zeitsignals kann bereits eine überraschend gute Sprachpausen-Detektion durchgeführt werden, die nur einen sehr geringen Rechenaufwand erfordert.

Eine Weiterbildung dieser Verfahrensvariante sieht vor, dass die drei Ausgangssignale  $\text{sam}(x)$ ,  $\text{mam}(x)$  und  $\text{lam}(x)$  zur Sprach-Pausen-Abschätzung einem neuronalen Netz zugeführt werden, das mit einer Vielzahl von Szenarien mit unterschiedlichen Eingangssignalen  $x$  trainiert wurde. Ein neuronales Netz kann vorteilhaft lineare und nichtlineare Zusammenhänge zwischen einer großen Menge von Eingabeparametern und den gewünschten Ausgabewerten abbilden. Eine Voraussetzung dafür ist, dass das neuronale Netz einmal mit einer hinreichenden Menge von Eingabewerten und

zugehörigen Ausgabewerten trainiert wurde. Daher eignen sich neuronale Netze besonders für die Aufgabe einer Sprachpausen-Detektion bei Anwesenheit von unterschiedlichen störenden Geräuschen.

Vorzugsweise wird zusätzlich zur Erkennung und Reduktion von Geräuschsignalen auch das Vorliegen von Echosignalen detektiert und/oder vorhergesagt und die entsprechenden Echosignale unterdrückt oder reduziert. Wenn in einem Telefonkanal zusätzlich zu Geräuschen auch Echos auftreten, so können diese in der Regel anhand einer vorab ermittelten Signallaufzeit  $\tau_E$  eines Echos sowie der vorab ermittelten Echokopplung ERL im Kanal und der Signalstärke ES, die das Echo im Rückkanal auslöst, vorhergesagt werden. Man kann diese Schätzung in der Weise durchführen, dass in Abhängigkeit vom ausgesendeten Sprachsignal und seiner momentanen Leistung die Größe der verzögert eintreffenden Echos abgeschätzt wird. Überschreitet das jeweils geschätzte Echosignal in bestimmten kurzen Zeitabschnitten einen vorgegebenen Schwellwert  $\text{thrs}$ , so wird dieses echobehaftete Signal vorzugsweise kurzzeitig zusätzlich, beispielsweise durch die oben erwähnte exponentielle Absenkung, auf einen Wert gedämpft, der für eine wesentliche Reduktion des Echosignals erforderlich ist. Im gleichen Sinn kann auch eine Kompander-Kennlinie kurzzeitig bei Echos in die Richtung größerer Eingangslautstärke verschoben und nach Abklingen der Echos wieder in ihre ursprüngliche Lage zurückgeführt werden.

Besonders bevorzugt ist eine Weiterbildung dieser Verfahrensvariante, bei der das Steuersignal  $a_0(k+1)$  laufend eingestellt wird gemäß  $a_0(k+1) = h(N, S, ES, \tau_E, ERL)$ , wobei  $h(N, S, ES, \tau_E, ERL)$  eine vorgegebene Funktion von  $N$ ,  $S$ , dem Nutzsignal  $ES$  in Gegenrichtung eines sprechenden TK-Partners,  $\tau_E$  eine konstante Verzögerungszeit des Echosignals und  $ERL$  eine Dämpfungskonstante der Amplitude des Echosignals ist.

Man kann vorteilhaft eine gehörrichtige Geräuschabsenkung mit einer unabhängig davon arbeitenden Echoreduktion verbinden. Das ist besonders

dann wichtig, wenn im Telefonkanal so gut wie kein Hintergrundgeräusch existiert, da dann keine Geräuschabsenkung wirksam wird, und somit auftretende Echos ungehindert zum Sprecher gelangen können.

Eine Trennung der Steuerung einer Geräuschreduktion von der einer Echoabsenkung ist zweckmäßig, da Geräusche und Echos unabhängig voneinander auftreten und auch in der Regel völlig unterschiedliche physikalische Ursachen haben. Man kann aber mathematisch eine generelle Reduktionsfunktion  $R$  angeben, die eine Absenkung von Signalpegeln sowohl für Geräusche als auch für Echos beschreibt:

$$R(S, N, ES, \tau_E, ERL, \text{thrs}) \sim g(S/N) \cdot d(ES, \tau_E, ERL, \text{thrs}),$$

wobei  $g(S/N)$  die oben beschriebene Geräuschreduktion und  $d(\dots)$  die unabhängig zusätzlich einsetzende Echoabsenkung bedeuten, wenn das geschätzte Echosignal den vorgegebenen Schwellwert  $\text{thrs}$  überschreitet.

Besonders vorteilhaft ist eine Verfahrensvariante, bei der während der Zeitdauer einer Echo-Reduktion zum Nutzsignal zusätzlich ein künstliches Geräuschsignal addiert wird.

Eine Geräuschabsenkung ist bei gleichbleibendem Geräuschpegel ebenfalls konstant. Eine plötzlich zusätzlich einsetzende Echoreduktion im Rhythmus der Sprache bedeutet auch eine Geräuschabsenkung (zumindest in dem kurzen Zeitabschnitt) im Sprachrhythmus. Dies führt zu einem gepulsten Hintergrundgeräusch, welches sich nicht natürlich anhört. Daher ist es vorteilhaft, in den Augenblicken einer zusätzlichen Echoreduktion ein synthetisches Geräusch eines geeigneten Rauschgenerators in der Größenordnung des normalen Hintergrundgeräusches zum bearbeiteten Signal hinzuzufügen. Damit soll ein möglichst gleichbleibendes Hintergrundgeräusch für den Hörer vermittelt werden.

Der Rauschgenerator kann so gestaltet werden, dass das künstliche Geräuschsignal eine psychoakustisch als angenehm empfundene akustische Signalsequenz (=comfort noise) umfasst.

Anstelle eines synthetischen Hintergrundgeräusches kann aber auch ein Abschnitt eines zuvor aufgezeichneten echten Hintergrundgeräusches in passender Stärke in den Echozeitabschnitten eingefügt werden. Das hinzugefügte Geräusch unterscheidet sich dann so gut wie überhaupt nicht vom vorherigen Geräusch und wird daher keine störende akustische Veränderung beim Hörer hervorrufen.

Das Hinzufügen von Geräuschen zur akustischen Verdeckung von Effekten sowie die Maßnahmen einer getrennten Behandlung von Geräuschen und Echos werden, wenn sie richtig aufeinander abgestimmt sind, einen besonders verständlichen und angenehmen Spracheindruck auch bei "schwieriger" Umgebung (Echos plus Geräusche) bewirken.

Besonders bevorzugt ist auch eine Variante des erfindungsgemäßen Verfahrens, bei der das zu übertragende Nutzsignal einer spektralen Subtraktion unterzogen wird. Der Vorteil einer spektralen Subtraktion mit nachgeschalteter Pegelabsenkung in den Sprachpausen besteht darin, dass zuerst mittels spektraler Subtraktion ein Teil der Störgeräusche aus dem Sprachsignal selbst eliminiert wird und erst danach die Sprachpausen in der beschriebenen Art von Geräuschen und Echos befreit werden. Insgesamt ergibt diese Kombination bei subjektiven Tests bessere Höreindrücke als nur eine einfache spektrale Subtraktion.

Eine weitere besonders vorteilhafte Variante des erfindungsgemäßen Verfahrens schließlich sieht vor, dass das zu übertragende Nutzsignal einer dem menschlichen Gehör angepassten spektralen Filterung unterzogen wird. Auch hier wird mit den Mitteln einer spektralen Subtraktion zunächst eine Schätzung von Geräuschen, von Sprache und den Echos durchgeführt, sodann

eine gehörrichtige Verdeckungsschwelle bestimmt und dann das gesamte Signal über ein passend eingestelltes Übertragungsfilter so bearbeitet, dass die Sprachanteile möglichst unverfälscht und die Echo- und Geräuschanteile möglichst weitgehend unterdrückt werden.

Eine Kombination mit der nachgeschalteten Pegelabsenkung in den Sprachpausen verbessert den Höreindruck weiter.

In den Rahmen der vorliegenden Erfindung fällt auch eine Servereinheit zur Unterstützung des oben beschriebenen erfindungsgemäßen Verfahrens sowie ein Computerprogramm zur Durchführung des Verfahrens. Das Verfahren kann sowohl als Hardwareschaltung, als auch in Form eines Computerprogramms realisiert werden. Heutzutage wird eine Software-Programmierung für leistungsstarke DSP's bevorzugt, da neue Erkenntnisse und Zusatzfunktionen leichter durch eine Veränderung der Software auf bestehender Hardwarebasis implementierbar sind. Verfahren können aber auch als Hardwarebausteine beispielsweise in TK-Endgeräten oder Telefonanlagen implementiert werden.

Weitere Vorteile der Erfindung ergeben sich aus der Beschreibung und der Zeichnung. Ebenso können die vorstehend genannten und die noch weiter aufgeführten Merkmale erfindungsgemäß jeweils einzeln für sich oder zu mehreren in beliebigen Kombinationen Verwendung finden. Die gezeigten und beschriebenen Ausführungsformen sind nicht als abschließende Aufzählung zu verstehen, sondern haben vielmehr beispielhaften Charakter für die Schilderung der Erfindung.

Die Erfindung ist in der Zeichnung dargestellt und wird anhand von Ausführungsbeispielen näher erläutert. Es zeigen:

Fig. 1            das Steuersignal  $a_0$  bei Vorliegen von Sprachsignalen, während einer Sprachpause und bei erneutem Einsetzen der Sprachsignale;

Fig. 2 ein Schema einer Anordnung zur gesteuerten Signalabsenkung;

Fig. 3a die Funktion  $g(S/N)$  in der linearen Näherung;

Fig. 3b die entsprechende Funktion  $g'(N/S)$ ;

Fig. 4a die Funktion  $g(S/N)$  als unsymmetrische Glockenkurve; und

Fig. 4b die entsprechende Funktion  $g'(N/S)$ .

Das in Fig. 1 als Funktion der Zeit  $t$  bzw. der Abtastzahl  $k$  dargestellte Steuersignal  $a_0$  wird während einer ersten Phase  $T1$ , in welcher Sprachsignale detektiert werden, auf einem Wert  $c_0 = 1$  gehalten. Während einer Sprachpause im Zeitabschnitt  $T2$  wird das Steuersignal  $a_0$  exponentiell auf einen kurz über 0 liegenden konstanten Wert  $c_2$  abgesenkt, um dann bei erneutem Einsetzen der Sprachsignale während einer Phase  $T3$  sprunghaft wieder auf den Wert  $c_0 = 1$  (oder eine andere, beliebig wählbare Konstante) heraufgesetzt zu werden. Dadurch wird während der Sprachphasen  $T1$ ,  $T3$  keine (oder bei anderen Beispielen nur eine geringe) Unterdrückung von Störsignalen im Gesamtsignal vorgenommen, so dass das Sprachsignal möglichst unverfälscht und ungehindert weitergeleitet wird. Während der Sprachpause in der Phase  $T2$  wird möglichst schnell (exponentiell) eine möglichst effektive Unterdrückung von Echos und Geräuschsignalen bewirkt, wobei aber im vorliegenden Beispiel nicht auf den Wert 0, sondern auf einen kleinen Restwert  $c_2$  abgesenkt wird, um nicht am anderen Ende den Eindruck einer "toten" Leitung zu erwecken. Beim Auftreten von Echos wird eine Absenkung auf einen Restwert  $c_3 < c_2$  vorgenommen.

In Fig.2 ist schematisch die Funktionsweise einer Anordnung zur Geräusch- und Echoreduktion entsprechend der oben genannten Reduktionsfunktion  $R(S, N, ES, \tau_E, ERL, \text{thrs})$  mit einem Sprach-Pausen-Detektor SPD dargestellt.

Für sämtliche in den Figuren 3a bis 4b dargestellten Kurven gilt, dass der Funktionswert  $g$  bzw.  $g'$  für den Fall  $S/N < 0$  dB, also bei extrem hohem Geräuschhintergrund, in einen konstanten Wert  $g_0$  der Geräuschreduktion von etwa 6 dB übergeht. Beginnend von  $S/N = 0$  dB wird mit zunehmender Verbesserung des Signal-zu-Rausch-Verhältnisses  $S/N$  eine erhöhte Geräuschreduktion vorgenommen, die ein Maximum  $g_{\max} \approx 25$  dB bei etwa  $S/N \approx 12$  dB erreicht. Mit weiter zunehmendem  $S/N$  sinkt der Grad der Geräuschreduktion schließlich gegen Null, um bei geringem Hintergrundgeräusch möglichst wenig Manipulationen im übertragenen Nutzsignal vorzunehmen.



### Zusammenfassung

Ein Verfahren zur Reduktion von Echo- und/oder Geräuschsignalen bei TK-Systemen für die Übertragung von akustischen Nutzsignalen, bei dem mittels Sprach-Pausen-Detektion ermittelt wird, wann eine Sprachpause vorliegt, wobei die gestörten Nutzsignale durch ein zeitabhängiges Steuersignal  $a_0(t)$  bzw. durch ein im Rhythmus einer Abtastrate  $f_T=1/T$  getaktetes Steuersignal  $a_0(k)$  verändert werden, ist dadurch gekennzeichnet, dass das Steuersignal  $a_0(k)$  so variiert wird, dass während des Vorliegens von Sprachsignalen im Nutzsignal die Amplitude des Steuersignals  $a_0(k)$  auf einen vorgegebenen konstanten Wert  $c_0$  gesetzt wird und mit Beginn einer Sprachpause die Amplitude des Steuersignals  $a_0(k)$  von einem Abtastwert zum nächsten gemäß der Rekursionsformel  $a_0(k+1) = a_0(k) \cdot \beta$  mit  $\beta < 1$  stetig abgesenkt wird, und dass nach dem Ende einer Sprachpause  $a_0(k) = c_0$  gesetzt wird. Damit kann unaufwendig und kostengünstig ohne großen Rechenaufwand und mit geringem Bedarf an Rechenspeicher und Datenspeicherplatz eine Echo- und Geräuschabsenkung bewirkt werden, die mit einfachen Mitteln einen für das menschliche Ohr möglichst angenehmen akustischen Gesamteindruck erzeugt, der je nach Geschmack an individuelle Bedürfnisse angepasst werden kann.

(Fig. 1)

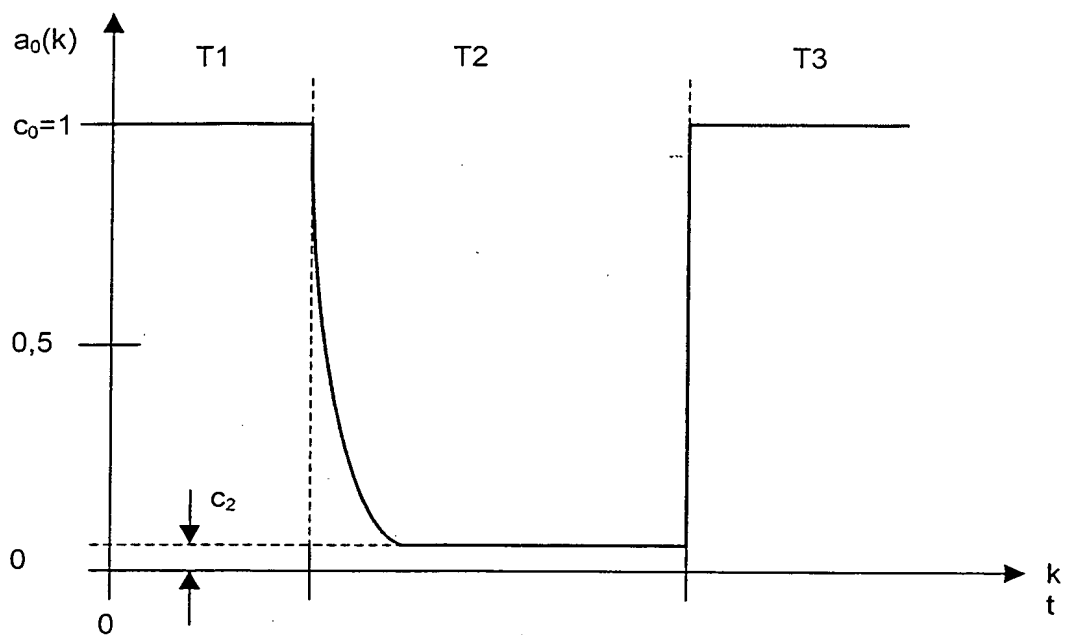


Fig. 1

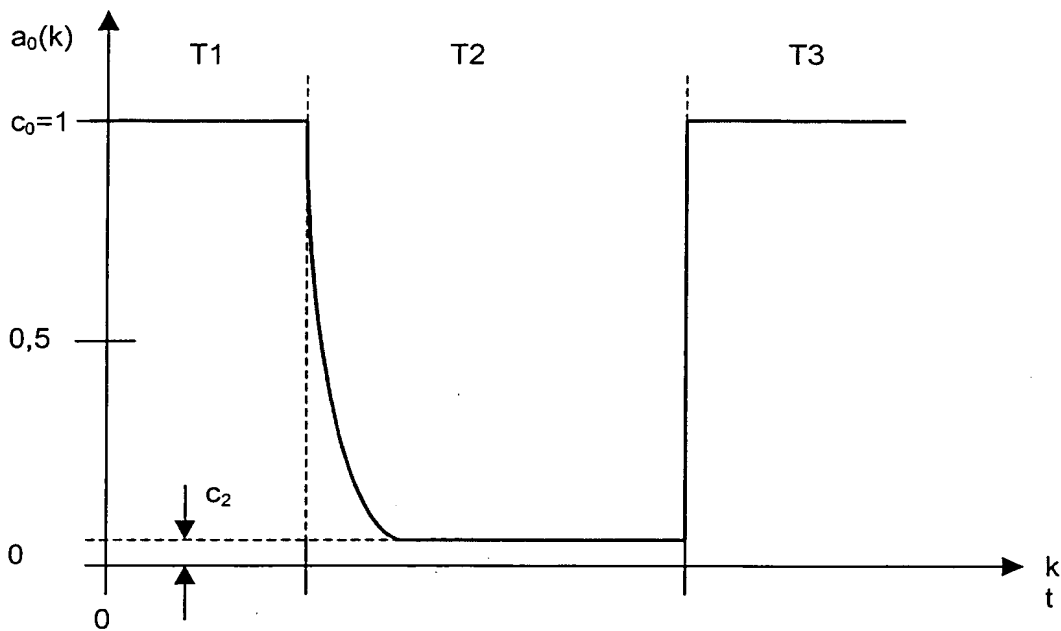


Fig. 1

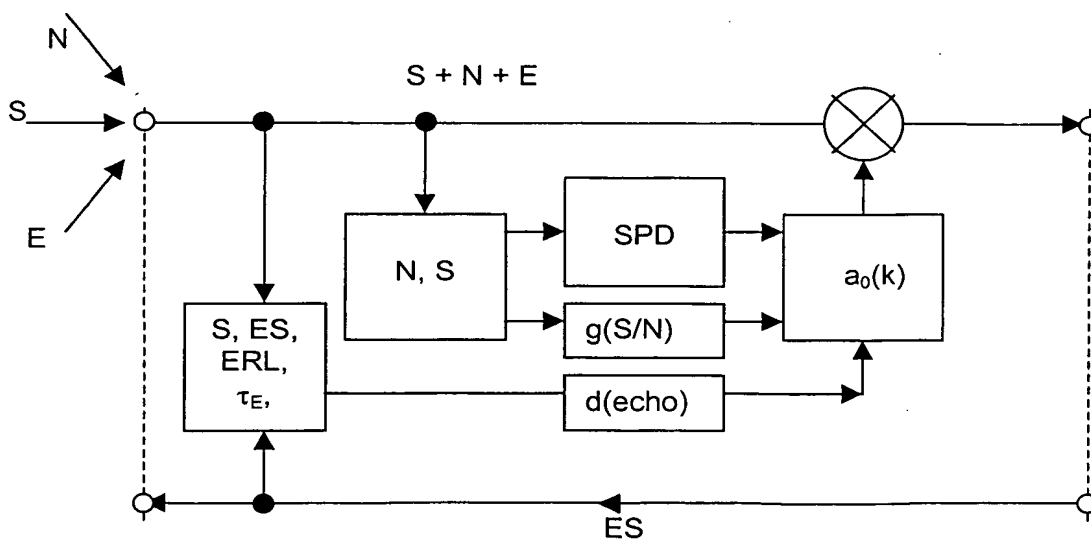


Fig. 2

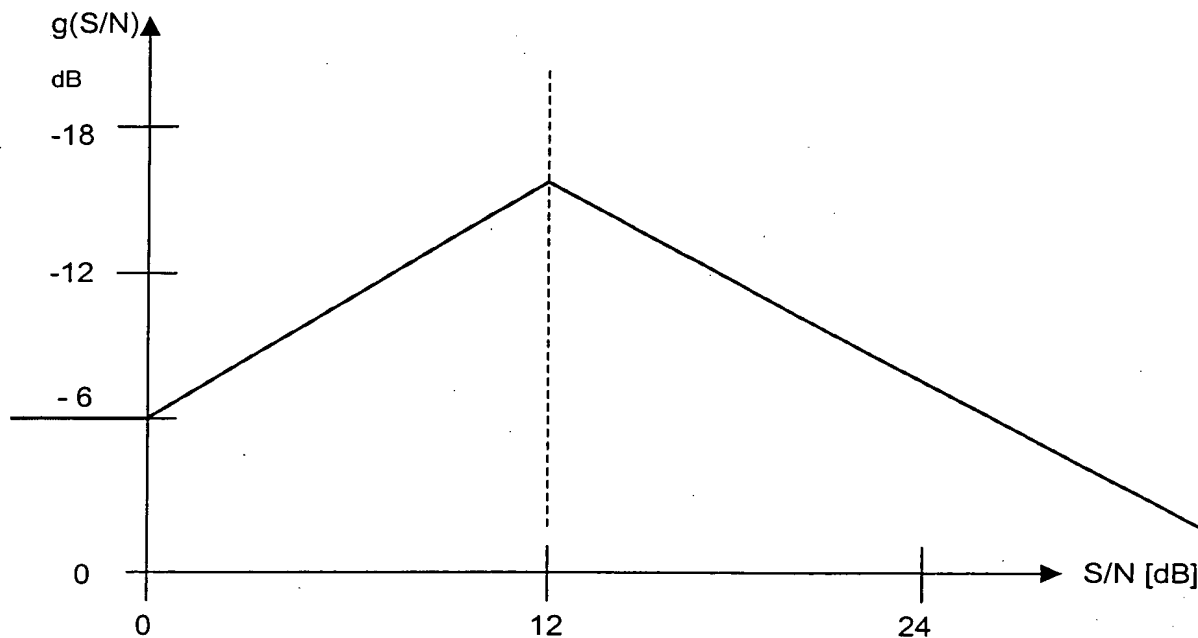


Fig. 3 a

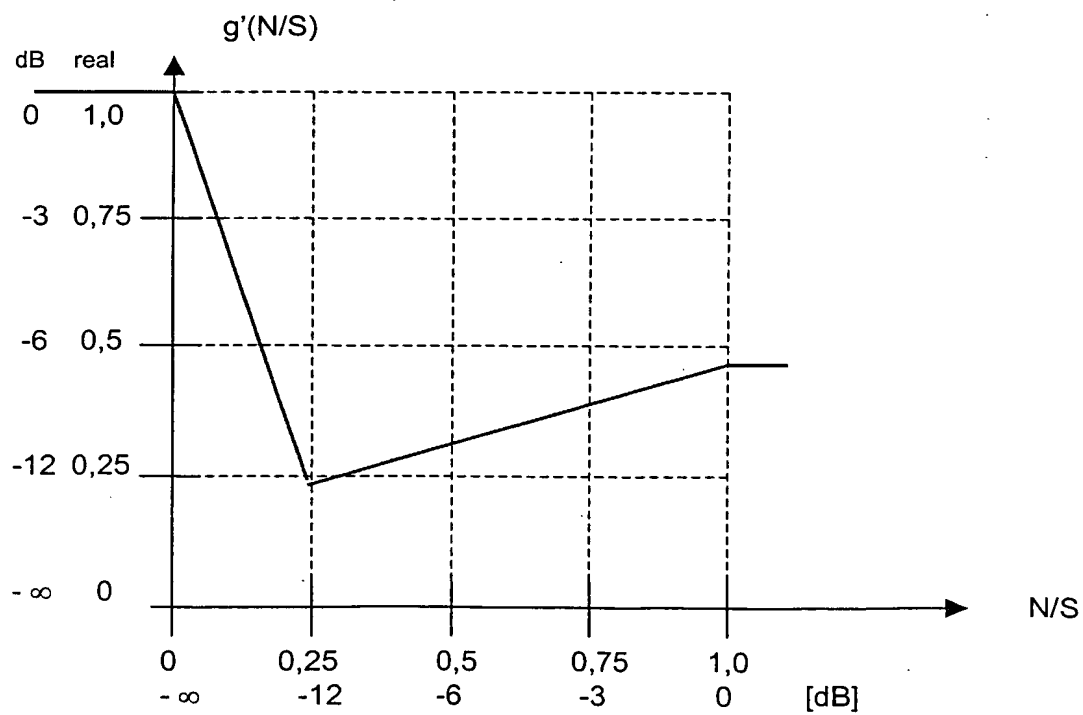


Fig. 3 b

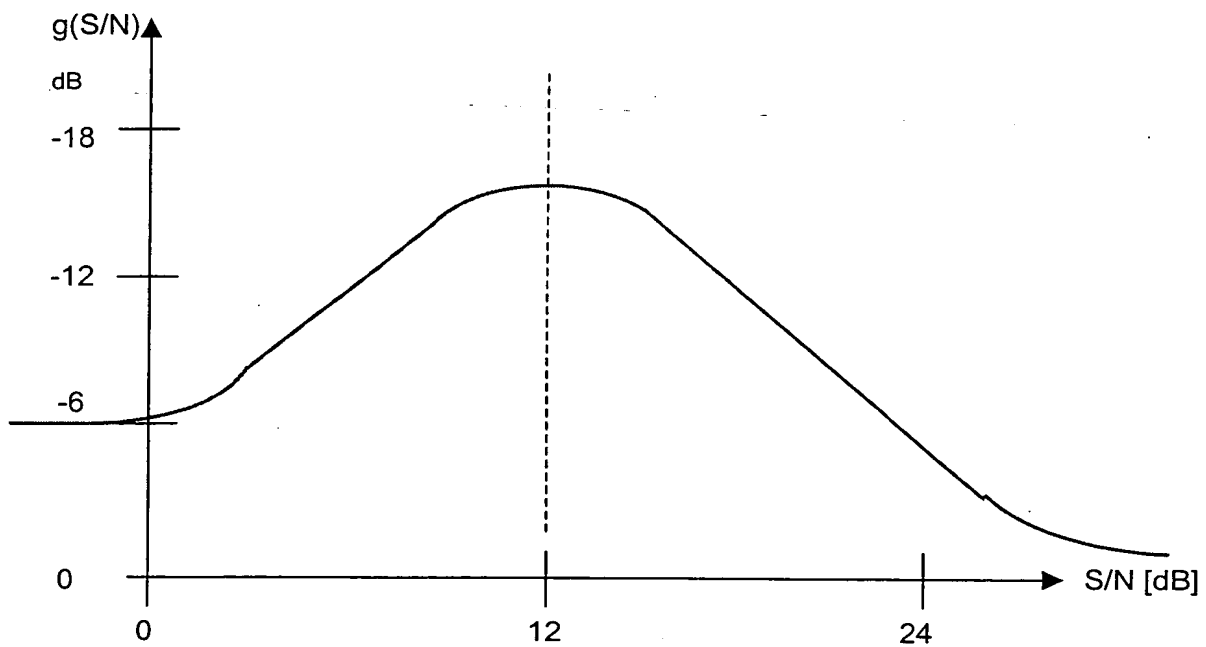


Fig. 4 a

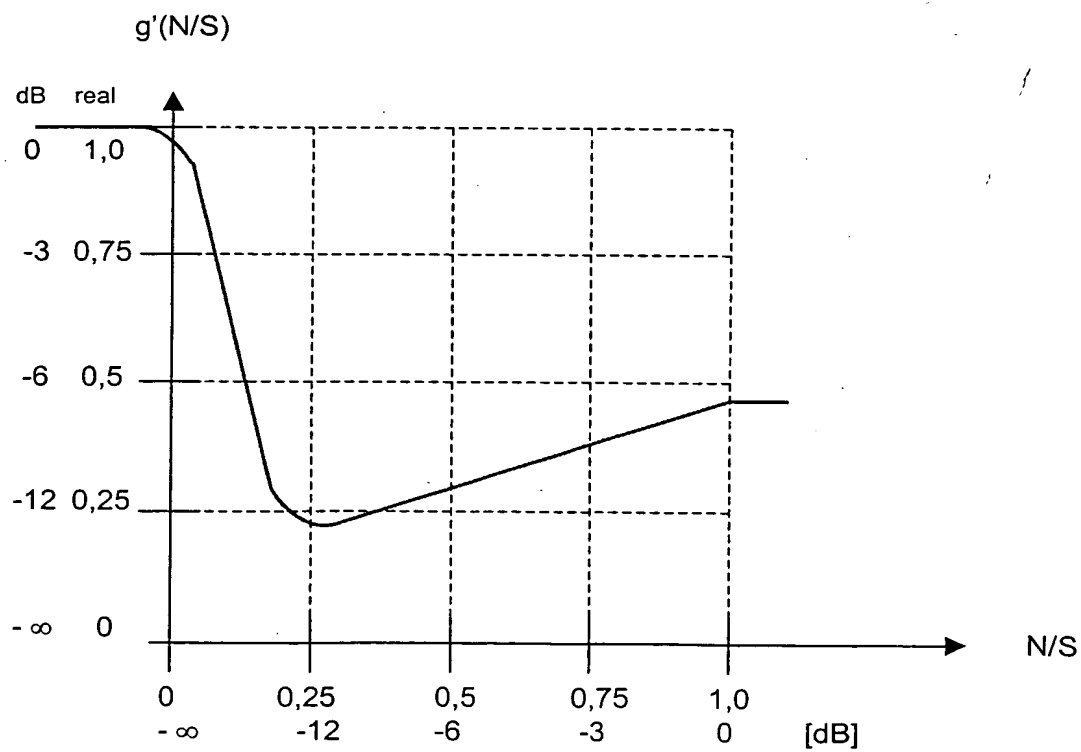


Fig. 4 b



Creation date: 11-02-2004  
Indexing Officer: KJONES4 - KIMBERLY JONES  
Team: OIPEBackFileScanning  
Dossier: 09716272

Legal Date: 07-11-2004

No.	Doccode	Number of pages
1	SRNT	2

Total number of pages: 2

Remarks:

Order of re-scan issued on .....